5.0 MONTE-CARLO MODULE

Monte-Carlo simulation is a statistical technique by which a quantity is calculated repeatedly, as many as thousands of times, using randomly selected parameter values for each calculation. The results approximate the full range of possible outcomes, and the likelihood of each. When Monte-Carlo simulation is applied to risk assessment, risk appears as a frequency distribution.

The Monte-Carlo simulation technique was developed during World War II and is named after the casinos in Monte Carlo, Monaco, where the primary attractions are games of chance. The random behavior in games of chance is similar to how Monte-Carlo simulation selects variable values at random to simulate a particular modeling scenario. When we roll a die, we know that either a 1, 2, 3, 4, 5, or 6 will come up, but we do not know *a priori* which value it will be for any particular roll. It is the same with the variables that have a known range of values but an uncertain value for any particular time or event.

This section presents the Monte-Carlo module of the EPACMTP model and describes how this probabilistic module is implemented for ground-water fate and transport analyses. The purpose of the Monte-Carlo module and a description of its operation are presented in Sections 5.1 and 5.2, respectively. The methods used to ensure that each model realization uses an internally consistent set of data are summarized in Section 5.3. Section 5.4 describes each of the distribution types that can be selected for each of the EPACMTP input parameters. Section 5.5 explains in detail how the regional, site-based Monte-Carlo methodology is implemented in EPACMTP; and Section 5.6 describes how to interpret the results of a Monte-Carlo modeling analysis. Finally, Section 5.7 summarizes the results of an analysis performed by EPA to determine the appropriate number of Monte-Carlo computer runs needed to achieve reliable results.

5.1 PURPOSE OF THE MONTE-CARLO MODULE

Application of the EPACMTP model for determining receptor well concentrations requires values for the various source-specific, chemical-specific, unsaturated-zone-specific and saturated-zone-specific model parameters. For many assessment purposes it is not appropriate to assign single values to all of these parameters. Rather, their values represent a probability distribution, reflecting both the range of variation that may be encountered at different waste sites around the country, as well as our uncertainty about the specific conditions at each site.

The Monte-Carlo module in EPACMTP makes it possible to incorporate uncertainty and variability in the values of parameters into the subsurface pathway modeling analysis, and to quantify the impact of parameter variability and uncertainty on expected receptor well concentrations. In particular, we use Monte-Carlo simulation to determine the likelihood, or probability, that the concentration of a constituent at the receptor well, and hence exposure and risk, will be above or below a certain value.

5.1.1 Treatment of Uncertainty and Variability

Variability arises from true heterogeneity in characteristics, such as rainfall at different locations in the United States. **Uncertainty** represents lack of knowledge about factors and processes, such as the effective hydraulic conductivity of the aquifer at a given waste management unit site or the nature of degradation mechanisms, that affect constituent fate and transport.

EPA classifies the major areas of uncertainty in risk assessments as parameter uncertainty, scenario uncertainty, and model uncertainty. Parameter uncertainty is the "uncertainty regarding some parameter" of the analysis. Scenario uncertainty is "uncertainty regarding missing or incomplete information needed to fully define exposure and dose." Model uncertainty is "uncertainty regarding gaps in scientific theory required to make predictions on the basis of causal inferences" (U.S. EPA, 1992).

The sources of parameter uncertainty are measurement errors, sampling errors, variability, and use of generic or surrogate data (U.S. EPA, 1992). In other words, many of the input parameters used to quantify contaminant fate and transport cannot be measured precisely and/or accurately.

The sources of scenario uncertainty include: estimation errors of operational periods, approximations of operational conditions, and disposal history of constituents in waste management units. Many of the operational conditions are so complex that the respective simplified approximations may not describe the true conditions precisely. In addition, the amount of data relating to operational conditions may not be adequate or may be subject to high degree of uncertainty.

The sources of model uncertainty are relationship errors and modeling errors (U.S. EPA, 1992). Models and their mathematical expressions are simplifications of reality that are used to approximate real-world conditions and processes and their relationships. Models do not include all parameters or equations necessary to express reality because of the inherent complexity of the natural environment and the lack of sufficient data to fully describe it. Consequently, models are based on various assumptions and simplifications and reflect an incomplete understanding of natural processes.

In the remainder of Section 5, we will use the term 'uncertainty' to cover both parameter variability and uncertainty. As explained above, strictly speaking, variability and uncertainty are different concepts. Variability describes parameters whose values are not constant in space and/or time; however, at least in principle, these parameter values can be measured or estimated and specified as a frequency distribution in the modeling input file. Uncertainty pertains to parameters, processes and relationships that we know or can model only approximately. In practice, we use probability distributions to describe both variability and uncertainty, and for the purpose of the EPACMTP Monte-Carlo module, we treat variability and uncertainty as equivalent.

The EPACMTP model accounts for the variability and uncertainty in environmental setting through the use of several linked databases: 1) a nationwide database of waste management unit sites and the environmental setting for each, 2) a database of the characteristics of each type of environmental setting (e.g. aquifer thickness and hydraulic conductivity), and 3) databases of climatic parameters (e.g., ground-water temperature, infiltration rate, and regional recharge rate). That is, through the use of these linked databases the EPACMTP model accounts for both the nationwide variability in environmental conditions and the uncertainty about these conditions at any given site. A fundamental underlying assumption in EPA's implementation of the EPACMTP Monte-Carlo module is that uncertainty in local conditions can be approximated using data that characterize the variability of sites across the United States.

In planning a Monte-Carlo modeling analysis, it is desirable to specifically address as much of the parameter variability and uncertainty as possible, either directly in the Monte-Carlo modeling process or through disaggregation of the data into discrete elements of the analysis. The use of a distribution of distances to the nearest downgradient receptor well accounts for spatial variability in concentrations around a WMU and uncertainty in receptor locations and is an example of doing this directly in the modeling process. The WMU site databases are an example of how disaggregation of the data can be used to address parameter uncertainty and variability. For a typical nationwide analysis conducted for regulatory purposes, a given waste stream may be disposed in a number of WMUs located all across the country. In modeling this scenario with EPACMTP, we account for the variability of WMU characteristics (such as area, depth, and operational life) by using large WMU site databases that were created by surveying a representative sample of the existing WMUs. Each record in the database represents one possible WMU site, and one record (or one set of correlated data representing an individual WMU) is selected for each model realization, such that at the end of the Monte-Carlo analysis the modeling results reflect the range of possible WMU characteristics.

5.2 MONTE-CARLO MODULE OPERATION

The Monte-Carlo method requires that for each input parameter, except constant and derived parameters, a probability distribution be provided. The method involves the repeated generation of pseudo-random values of the uncertain input variable(s) (drawn from the known distribution and within the range of any imposed bounds). The EPACMTP model is executed for each set of randomly generated model parameters and the corresponding receptor well exposure concentration is computed and stored. Each set of input values and corresponding receptor well concentration is termed a ***realization***.

A typical Monte-Carlo simulation can involve thousands of realizations. At the conclusion of the Monte-Carlo simulation, the realizations are statistically analyzed to yield a cumulative probability distribution of the receptor well exposure concentration. The various steps involved in the application of the Monte-Carlo simulation technique are:

(1) Select representative probability distribution functions for the relevant input variables.

(2) Generate random values from the distributions selected in (1). These values represent a possible set of values (a *realization)* for the input variables.

(3) Run EPACMTP with these input values. Store the resulting receptor well exposure concentration.

(4) Repeat Steps (2) and (3) for a specified number of times.

(5) Statistically analyze the computed receptor well concentrations to develop a cumulative probability distribution of either the receptor well concentration or a *Dilution Attenuation Factor* (DAF); the DAF is defined as the ratio of the initial leachate concentration to the receptor well concentration. In other words, it represents the reduction in constituent concentration that occurs before the leachate reaches the well.

The EPACMTP user performs Step 1 during the creation of the modeling input file(s), and the Monte-Carlo module of EPACMTP performs Steps 2 through 4. Step 5 is typically performed as a post-processing step using a spreadsheet or a utility program.

A simplified flow chart that illustrates the linking of the Monte-Carlo module to the simulation modules of the EPACMTP composite model is presented in Figure 5.1. The model input data are read first, followed by the generation of the random numbers. The generated random and/or derived parameter values are then assigned to the model variables. Following this, the contaminant transport fate and transport simulation is performed. The result is given in terms of the predicted contaminant concentration in a downgradient ground-water receptor well. The generation of random parameter values and fate and transport simulation is repeated as many times as necessary to accurately determine the probability distribution of receptor well concentrations.

5.3 ENSURING INTERNALLY CONSISTENT DATA SETS

As discussed in Section 5.1, a ground-water modeler needs input values for many waste, chemical, and subsurface parameters in order to perform a ground-water pathway analysis using the EPACMTP model.

Inherent in the Monte-Carlo process is that parameter values are drawn from multiple data sources, and then combined in each realization of the modeling process. Because the parameter values are drawn randomly from their individual probability distributions, it is possible that parameters are combined in ways that are physically infeasible and that violate the validity of the EPACMTP flow and transport model. The Monte-Carlo module of EPACMTP incorporates three main methods to eliminate or reduce these occurrences as much as possible:

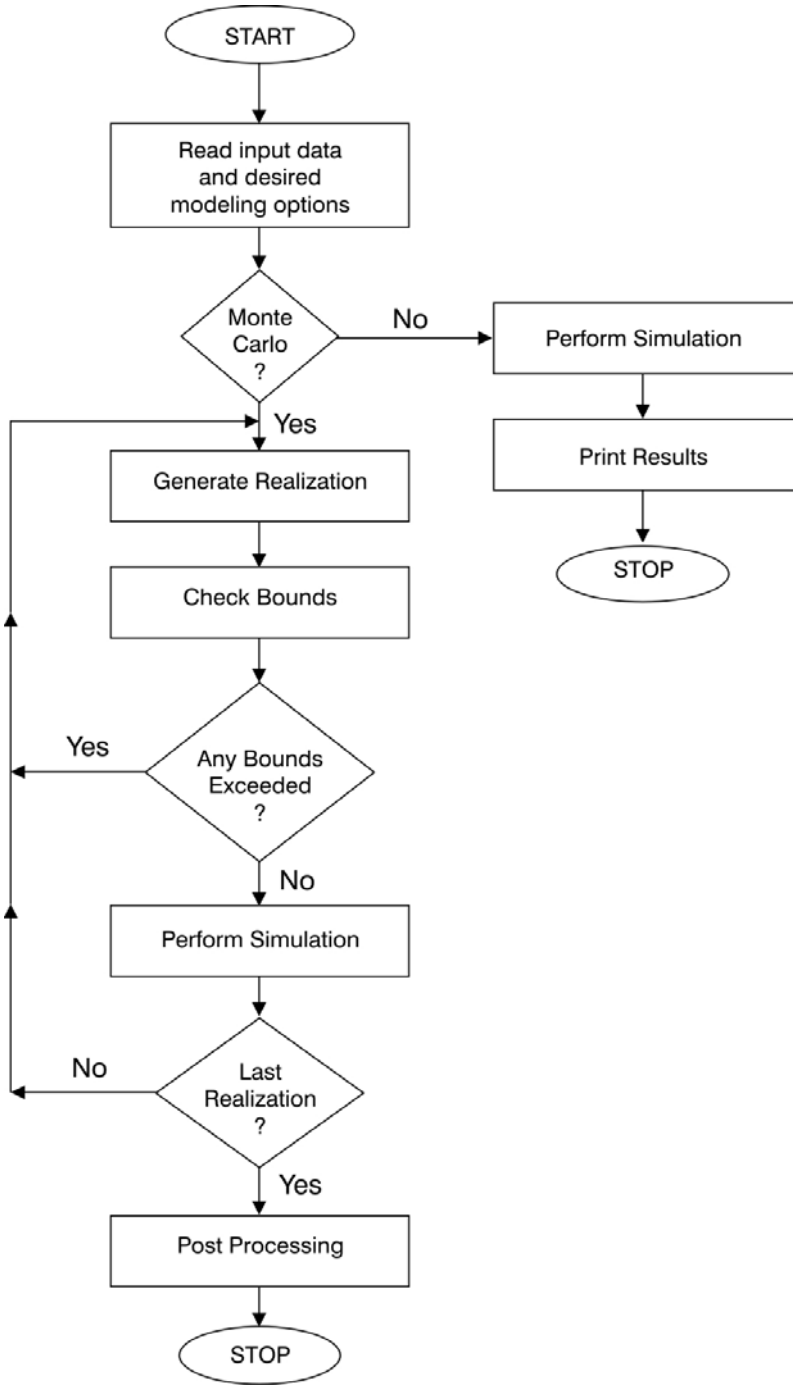

Figure 5.1 Flow chart of EPACMTP for a Monte-Carlo Problem.

- Impose upper and lower limits on parameters that are randomly chosen from defined distributions or internally calculated by the model (see Sections 5.3.1 and 5.4.11);

- Perform a ground-water table elevation screening procedure to ensure that the generated parameter values for a given Monte-Carlo realization do not result in a physically implausible scenario with respect to the elevation of the ground surface, the water table, and/or the top of an impoundment (see Sections 4.3.6 and 5.3.2); and

- Use of the regional site-based methodology which links together several correlated data sets (see Sections 5.3.3 and 5.5).

5.3.1 Upper and Lower Limits

As a relatively simple measure, upper or lower limits are specified on the values of individual Monte-Carlo parameters to ensure that their randomly generated values are within physically realistic limits. We also specified upper and lower limits on secondary parameters whose values are calculated (derived) internally in the Monte-Carlo module as functions of the primary EPACMTP input parameters, (see the *EPACMTP Parameters/Data Background Document* (U.S. EPA, 2003)).

5.3.2 Screening Procedures

In addition to the enforcement of upper and lower limits for randomly generated and derived input values, EPACMTP also automatically performs a set of screening procedures to ensure that the conceptual model remains physically plausible. These screening procedures are summarized below; additional details, mathematical formulations, and flow charts of the screening process are presented Section 4.3.6.

The ground-water elevation screening procedure is used to ensure that the generated parameter values for a given Monte-Carlo realization do not result in a physically implausible scenario with respect to the elevation of the ground surface, the water table, and/or the top of an impoundment. Physically, a rise of the water table above the ground surface would indicate the WMU is located in a swamp. The reason for implementing this type of screening is that the EPACMTP Monte-Carlo module may generate unrealistically high values for infiltration and recharge at a site with shallow depth to ground water and low hydraulic conductivity of the aquifer. And a rise in the water table above the height of the waste water in a surface impoundment would mean that ground-water transport would tend to be from the aquifer into the impoundment rather than from the impoundment into the aquifer; it is unlikely that an impoundment like this would be constructed to manage waste water.

For a given Monte-Carlo realization for landfills, waste piles, and land application units, the four correlated hydrogeological parameters, infiltration rate through the WMU, and ambient regional recharge rate are generated. Then the EPACMTP model calculates the estimated water table mounding that would result from the selected combination of parameter values. The combination of parameters is accepted if the calculated maximum water table elevation (the ground-water ‘mound’)

remains below the ground surface elevation at the site. If the criterion is not satisfied, the selected parameter values for the realization are rejected and a new data set is selected from the appropriate distributions.

For surface impoundments, there are two additional considerations in the screening process. In a typical Monte-Carlo modeling analysis for a surface impoundment, a site is selected from the surface impoundment WMU data base for each realization. The unit-specific parameters, including ponding depth and base depth below ground surface, are retrieved from the data base. The four correlated hydrogeologic parameters are then selected from the hydrogeologic data base, based on the hydrogeologic environment at that WMU location. The EPACMTP model then executes the probabilistic screener using these values for the base depth and water table elevation. If the elevation of the waste water surface in the impoundment is below the water table, that set of parameter values fails the screening process. In this case, the selected parameter values for that realization will be rejected and a new data set selected from the appropriate distributions.

If the base of the unit is located above the water table, the unit is said to be hydraulically separated from the water table. However, in this case, it is necessary to ensure that the calculated infiltration rate does not exceed the maximum feasible infiltration rate; that is, the maximum rate that does not cause the crest of the local ground-water mound to be higher than the base of the surface impoundment. This limitation allows us to determine a conservative infiltration rate that is based on the free-drainage condition at the base of the surface impoundment. If the maximum feasible infiltration rate (I_{max}) is exceeded, the EPACMTP model will set the infiltration rate to this maximum value.

For a surface impoundment, once these limits on the derived infiltration have been imposed, a check to ensure that any ground-water mounding does not result in a rise of the water table above the ground surface is performed in the same manner as for other types of WMUs.

5.3.3 Regional Site-Based Approach

The regional site-based approach is the third method incorporated into the EPACMTP model to reduce the likelihood that a physically infeasible set of environmental data will be generated. This modeling approach and the correlated data sets are summarized below, and additional details are presented Section 5.5.

The main advantage of this regional site-based approach over a strictly nationwide methodology is that it is based on correlated data sets compiled at actual waste sites around the country that are linked to databases of climatic and hydrogeologic parameters through the use of climate and hydrogeologic indices. Using these correlated and linked databases, the regional site-based approach can, for each Monte-Carlo realization, generate a random, yet internally consistent, set of the required site-specific values without requiring the exhaustive sampling that would be required to actually gather these data from waste sites around the country.

Fundamentally, the approach used for a site-based Monte-Carlo analysis consists of conducting the modeling analysis for the waste sites in the Subtitle D survey on the assumption that these sites are an adequate representation of the universe of possible waste sites in the U.S. For each Monte-Carlo realization, EPACMTP selects a site, at random, from the Subtitle D survey data set. The corresponding climatic and hydrogeologic indices and the generated soil type are then used by the model to generate random, but internally consistent, sets of values for the climatic, soil and aquifer parameters. Thus, the use of the regional site-based methodology which links together several correlated data sets serves to reduce the probability that the generated data set contains a physically infeasible, unrealistic, or highly unlikely set of parameter values. Additional details about the data sources and implementation of the regional site-based modeling method are presented in Section 5.5.

5.4 METHODOLOGY FOR GENERATING INPUT VALUES ACCORDING TO SPECIFIED DISTRIBUTIONS

Variables that are treated as random must be assigned one of the thirteen probability distribution types that are available in EPACMTP. The distribution types and their corresponding EPACMTP distribution type codes are listed in Table 5.1. The default distribution type for each Monte-Carlo variable is discussed in the *EPACMTP Parameters/Data Background Document* (U.S. EPA, 2003).

The first step in generating a parameter value from a specified distribution involves generating a uniformly distributed random number between zero and one, designated as U[0,1] (hereafter referred to as a uniform random number). The Monte-Carlo module uses standard FORTRAN pseudo-random number generation routines (provided in the FORTRAN compiler software) to generate this uniform random number. This uniform random number generator is initialized using a seed value. EPACMTP uses a constant seed value which means that when a EPACMTP Monte-Carlo simulation is repeated with the same data input file, the model will reproduce exactly the same results.

The second step consists of using this uniform random number in conjunction with the probability distribution specified in the input file to generate an appropriate value for the given EPACMTP input parameter; this process is referred to as a transformation and is more fully explained for each distribution type in Sections 5.4.2 through 5.4.9. More specifically, the Monte-Carlo module of EPACMTP uses FORTRAN software routines (documented by McGrath and Irving, 1973) to perform these parameter transformations – that is, to generate a random value from a specified distribution using a uniform random number.

These two steps are then repeated for each input parameter specified in the input file as a distribution of values until a complete set of modeling data is generated.

5.4.1 Constant

Constant parameters are set to a fixed value during the Monte-Carlo simulation.

Table 5.1 Probability distributions and their associated codes available for use in Monte-Carlo module of EPACMTP.

Distribution Type	EPACMTP Distribution Type Code
Constant	0
Normal	1
Lognormal	2
Exponential	3
Uniform	4
Log 10 Uniform	5
Empirical	6
Johnson SB	7
Gelhar Empirical[a]	8
Area Transformation[b]	9
Vertical Well Position[c]	12
Site-based[d]	99
Derived Variable	-1

[a]Gelhar's distribution applies only to saturated-zone dispersivities (Gelhar et al., 1992).

[b]This distribution applies to municipal Subtitle D landfill areas only.

[c]This distribution applies to the vertical position of the receptor well below the water table only; it is used to specify that the receptor well is located at a fixed depth below the water table.

[d]This distribution applies to parameters that are read directly from an ancillary data file that contains waste locations, volume and area, and corresponding regional climatic and hydrogeological parameters for the site-based Monte-Carlo analysis.

5.4.2 Normal Distribution

The normal, or Gaussian, distribution is given by:

$$f(x^*) = \frac{1}{\sigma_N\sqrt{2\pi}}\exp[-0.5(\frac{x^* - \mu_N}{\sigma_N})^2] \qquad (5.1)$$

where

x^* = random variable (normally distributed)
$f(x^*)$ = probability density function of x^*
σ_N = standard deviation of normal distribution, and
μ_N = mean of normal distribution

Normally distributed random variables are generated from U[0,1] values using the ANRMRN function in McGrath and Irving (1973). This function transforms a U[0,1] random value into a normally distributed value with mean of zero and unit standard deviation N[0,1]. The random parameter value for a normal distribution with mean μ and standard deviation σ is then given by

$$x^* = \sigma_N N[0,1] + \mu_N \quad (5.2)$$

where

x^*	=	random variable (normally distributed)
σ_N	=	standard deviation of normal distribution
$N[0,1]$	=	normally distributed value with mean of zero and standard deviation of one
μ_N	=	mean of normal distribution

5.4.3 Lognormal Distribution

A variable has a lognormal distribution if it is converted to a normal distribution by taking the natural log of its value(s). The lognormal distribution is given by:

$$Y_{LN} = \exp(x^*) \quad (5.3)$$

where

Y_{LN}	=	lognormally distributed random variable
x^*	=	normally distributed random variable

5.4.4 Exponential Distribution

A variable has a exponential distribution if it is converted to a normal distribution by taking the exponential of its value(s). The exponential distribution is given by:

$$Y_{exp} = ln(x^*) \quad (5.4)$$

where

Y_{exp}	=	exponentially distributed random variable
x^*	=	normally distributed random variable

5.4.5 Uniform Distribution

In a uniform distribution, each value has an equal (uniform) probability of occurrence. The user must specify the upper and lower bounds for the distribution. Uniformly distributed variables are generated from:

$$Y_u = A_{YU} + (B_{YU}-A_{YU})\ U[0,1] \quad (5.5)$$

where

Y_u	=	uniform random variable
A_{YU}	=	lower bound for Y_u
B_{YU}	=	upper bound for Y_u
$U[0,1]$	=	uniform random number between zero and one

5.4.6 Log_{10} Uniform Distribution

A variable has a $\log_{10}$ uniform distribution if it is converted to a uniform distribution by taking the logarithm to the base 10 of its value(s). The user must specify upper and lower bounds for the distribution. The upper and lower bounds are specified in the same units as the actual EPACMTP variable of interest, but the bounds are internally converted to $\log_{10}$ values. The $\log_{10}$ uniform distribution is given by:

$$Y_{LU} = 10^{(A''+(B''-A'')U[0,1])} \quad (5.6)$$

where

Y_{LU}	=	$\log_{10}$ uniform random variable
A''	=	$Log_{10}(A_{YLU})$
A_{YLU}	=	lower bound for Y_{LU}
B''	=	$Log_{10}(B_{YLU})$
B_{YLU}	=	upper bound for Y_{LU}
$U[0,1]$	=	uniform random number between zero and one

5.4.7 Empirical Distribution

The empirical distribution is the most flexible probability distribution allowed in EPACMTP. Whereas other distributions assume that the probability of the actual modeling parameter of interest can be described by a particular type of mathematical equation, the empirical distribution does not make any assumptions about the underlying probability distribution of the data. The empirical distribution is simply a tabulation of parameter values and their corresponding frequency of occurrence. The empirical distribution is therefore well suited to empirically measured data, especially when there are relatively few measured data points. For empirical distributions, the user must provide a table of (measured) data values and their corresponding cumulative frequency of occurrence. The frequency is normalized from zero to one. EPACMTP generates random values for an empirical distribution as follows: First, it generates a uniform random number, U[0,1], representing the normalized cumulative frequency for the empirical parameter of interest. Next, it reads the corresponding parameter value off the table of data values and their frequency. Linear interpolation between the next lowest and next highest data values is used when the randomly generated probability value does not exactly match any of the tabulated frequency values.

For example, let us consider a simple empirical distribution with 4 data points, listed in increasing order as shown below in Table 5.2.

Table 5.2 Example Empirical Distribution

Relative Cumulative Frequency	Value
F1	A_E
F2	B_E
F3	C_E
F4	D_E

By definition, the relative cumulative frequency for the lowest value in the table (A_E) is zero (*F1* = 0); for the highest value in the table (D_E), it is one (*F4* = 1.0). To generate a random value for the parameter, EPACMTP first generates a random probability, R_n:

$$R_n = U[0,1] \tag{5.7a}$$

where

R_n = generated random number which corresponds to the cumulative probability of Y
U[0,1] = a uniform random number between zero and one

The code then performs a table look-up to find the next lower and next higher value in the frequency column of the table, and calculates the corresponding value for the parameter of interest by linear interpolation. Assume for the current example that the probability, R_n, lies in between *F3* and *F4*. The corresponding parameter value is then calculated as:

$$Y_E = (1 - \frac{R_n - F3}{F4 - F3})C_E + (\frac{F4 - R_n}{F4 - F3})D_E \tag{5.7b}$$

where

Y_E = the random variable with empirical distribution
R_n = generated random number which corresponds to the cumulative probability of Y_E
F_3, F_4 = cumulative probabilities for C_E and D_E, respectively
C_E = parameter value whose cumulative probability is *F3*
D_E = parameter value whose cumulative probability is *F4*

5.4.8 Johnson SB Distribution

The Johnson SB distribution (McGrath and Irving, 1973) represents a special transformation applied to a random variable such that the transformed variable is normally distributed. The Johnson SB distribution is given by:

$$Y_{JSB} = \frac{(A_{YJ} + B_{YJ}e^{x^*})}{(1 + e^{x^*})} \quad (5.8)$$

where

Y_{JSB}	=	random variable with Johnson SB distribution
A_{YJ}	=	lower bound for Y_{JSB}
B_{YJ}	=	upper bound for Y_{JSB}
x^*	=	normally distributed random variable

5.4.9 Special Distributions

In addition to the general distributions that can be used for any EPACMTP input parameter, the model handles a number of special distributions, each of which is unique to a particular EPACMTP model parameter. These special distributions are presented in this section.

5.4.9.1 Gelhar Distribution for Aquifer Dispersivity

The transport of the contaminant plume in the saturated zone is controlled by two mechanisms: advection and dispersion; the EPACMTP saturated-zone flow module simulates both of these mechanisms. Dispersion is the phenomenon by which a contaminant plume in flowing ground water is mixed with uncontaminated water and becomes reduced in concentration at the perimeter of the plume. Not all of a contaminant plume is traveling at the same velocity due to differences in pore size and flow path length and friction along pore walls, resulting in mixing along the flow path which decreases solute concentrations.

The model computes the longitudinal (along the flow path, or in the x-direction), horizontal transverse (perpendicular to the flow path, or in the y-direction), and vertical (in the z-direction) dispersion coefficients as the product of the seepage velocity and longitudinal (α_L), transverse (α_T) and vertical (α_V) dispersivities. A literature review indicated the absence of a generally accepted theory to describe dispersivities, although a strong dependence on scale has been noted (Gelhar et al., 1985; Gelhar et al., 1992). In a typical Monte-Carlo modeling analysis performed with EPACMTP, the longitudinal dispersivity is represented through a probabilistic formulation and the horizontal transverse and vertical dispersivities are then calculated from the longitudinal dispersivity, as summarized below; further details are given in Section 5.3.8 of the *EPACMTP Parameters/Data Background Document* (U.S. EPA, 2003).

In the absence of user-specified values or distributions, the longitudinal dispersivity is represented through a probabilistic formulation that is scaled based on well distance (Gelhar, 1986, personal communication), and the horizontal transverse and vertical dispersivities are then calculated from the longitudinal dispersivity. By default, the transverse (α_T) and vertical (α_V) dispersivities are calculated as a fraction of the longitudinal dispersivity. The default values of the ratio of the longitudinal to the transverse dispersivity, α_L/α_T, and the ratio of the longitudinal to the vertical dispersivity, α_L/α_V, are 8 and 160, respectively. The rationale for these default values are presented in Section 4.4.3.2.

5.4.9.2 Vertical Well Intake Point Depth

The depth of the intake point below the water table is the depth at which the model calculates the resulting ground-water concentration. Unlike most wells in the real world that have a screened interval of several feet or more, the simulated receptor well in EPACMTP has an intake that is a single point in space, as if the well consisted of a solid casing that was open at the bottom. In this case, the intake point would be the same as the depth of the well (or ZWELL). This depth is measured from the water table, not from the ground surface. For a Monte-Carlo analysis, there are several options for determining the depth of the well intake point that are implemented through the use of different EPACMTP distribution type codes.

The default option is to model the vertical position of the well as being uniformly distributed between the water table ($z_{rw}^* = 0$) and the saturated aquifer thickness ($z_{rw}^* = B$). This option is selected by specifying the z_{rw}^*-position as a uniform distribution (EPACMTP distribution type code 4, Table 5.1) with lower and upper limits of 0.0 and 1.0. EPACMTP will multiply this uniformly generated value by the saturated-zone thickness to yield the actual receptor well depth below the water table for each Monte-Carlo iteration.

Alternatively, if the upper limit is greater than 1, the vertical position of the receptor well is modeled as being uniformly distributed between these two limits. If the computed depth is greater than the saturated thickness, a new well position and/or a new depth are generated.

As a second option, data on the depth of receptor wells obtained from Agency surveys can be used directly in the model as an empirical distribution. The data values range from 15 ft (4.5 m) to 301 ft (90.9 m). If the generated value for the vertical position of the receptor well intake point exceeds the saturated thickness of the aquifer or if it is less than the depth to the saturated zone, a new well position is generated.

As a third option, the well position may be fixed at a constant depth. In this case, a EPACMTP distribution type code of either 12 or 0 (see Table 5.1) can be used in the input file. Each of these codes refers to a constant depth (measured in meters) for the well intake.

For the first two options, the vertical position of the receptor well can also be constrained to lie within the approximate vertical penetration depth of the

contaminant plume emanating from the waste unit, as defined by Equation (6.14) in Section 6 of the *EPACMTP Parameters/Data Background Document* (U.S. EPA, 2003).

5.4.10 Derived Parameters

In EPACMTP, a derived parameter is the one whose value is calculated directly from one or more other EPACMTP variables. Usually one or more of these variables has a probability distribution, so that the value of the derived variable also follows a frequency distribution. The relationships that are used to calculate values for derived variables represent direct physical relationships. An example is the relationship between aquifer porosity and bulk density. If we know porosity, it is possible to estimate bulk density and vice versa.

The individual parameters that are treated as derived parameters in EPACMTP, and their dependence on other EPACMTP variables are presented in the *EPACMTP Parameters/Data Background Document* (U.S. EPA, 2003).

5.4.11 Parameter Upper and Lower Bounds

Upper and lower bounds can be specified for each of the EPACMTP Monte-Carlo parameters. A number of the probability distributions in EPACMTP, including the uniform, Log_{10} uniform, empirical and the Johnson SB distribution already incorporate upper and lower bound values in the algorithm itself, so that the generated values for these distributions always fall within the allowable range. In the case of the normal and lognormal distribution, EPACMTP first generates a random value, compares it to the upper and lower bounds that are specified for that parameter, and if necessary regenerates new values until an acceptable value is obtained. EPACMTP follows a modified procedure for derived parameters. Because derived parameters reflect physical dependencies on other EPACMTP model parameters, it is not appropriate to simply modify the value of a derived parameter. Derived parameters may depend on more than one other EPACMTP parameter, which in turn may be related to more than one derived parameter. To deal with these multiple dependencies, EPACMTP regenerates the entire set of Monte-Carlo parameters in that realization until all parameter bounds are satisfied.

5.5 MONTE-CARLO METHODOLOGY FOR REGIONAL SITE-BASED, CORRELATED DISTRIBUTIONS

In reality, many of the site characteristics that control contaminant fate and transport are correlated with one another. For instance, climatic characteristics that drive infiltration and recharge, as well as soil and aquifer properties, are a function of a site's location. Except for derived parameters, the Monte-Carlo methodology and EPACMTP distribution types described in Section 5.4 treat each model parameter as independent. Parameter upper and lower bounds ensure that the value of each parameter is within a reasonable range, but they do not guarantee that the combination of parameter values that is randomly generated necessarily represents realistic site conditions. For instance, many of the Monte-Carlo input distributions that have been developed for EPACMTP (see the *Parameters/Data Background*

Document (U.S. EPA, 2003)) reflect nationwide variability. Random sampling from these individual distributions may result, for instance, in combining recharge rates from the arid southwestern United States, with ground-water depths that are typical of Florida. For situations in which the appropriately correlated parameter data sets are available, EPACMTP's regional, site-based, Monte-Carlo capability overcomes the above-mentioned limitation.

5.5.1 Description of Regional Site-Based Approach

The Monte-Carlo methodology implemented in EPACMTP is called 'regional site-based' because waste site databases are linked by each site's geographic location and underlying aquifer type to regional databases of climatic and subsurface parameters, respectively. In this way, the regional site-based approach attempts to approximate the ideal situation where we have a complete set of the site-specific input data required to run the EPACMTP model for each waste site in a statistically valid subset of the universe of waste management units in the United States.

In order to implement this site-based approach, the Agency has assembled a regional, site-based modeling database for each of the four types of waste management units that are typically modeled with EPACMTP (landfill, waste pile, surface impoundment, and land application unit). Additional details about the data included in these databases (and the corresponding data sources) are provided in the *EPACMTP Parameters/Data Background Document* (U.S. EPA, 2003). The remainder of this section briefly explains how the regional, site-based modeling approach is implemented.

The regional, site-based modeling procedure is based on an empirical distribution of waste sites, which can be envisioned as a list of sites; each record in the list corresponds to an actual waste site located somewhere in the U.S. For each record in the list, the site-related characteristics corresponding to that site are provided, and the value for each characteristic is specified as either as a single value or as a distribution or range of values. For instance, EPACMTP can handle either a specific value for the depth to ground water or a distribution of values, in case the specific value at that waste site is uncertain. However, in the site-based procedure, not all of the EPACMTP input parameters need to be specified as site-related. When inputs are not specified as site-related, the parameters are specified as one of the probability distribution types presented in Section 5.4. As an example, the receptor well location is often a non-site-related input parameter. Even if data on specific receptor well locations downgradient from waste sites are available, these locations may change in the future, thus, the user may want to consider a range or probability distribution of receptor well locations in conducting a risk assessment.

The data sources for the regional site-based methodology that are typically used to conduct a Monte-Carlo modeling analysis with EPACMTP include: 1) the Hydrogeologic DataBase for Modeling (HGDB) (Newell et al., 1989; U.S. EPA, 1997), developed from a survey of hydrogeologic parameters for actual hazardous waste sites in the United States; 2) the infiltration and recharge analysis performed for 102 U.S. climatic centers using the HELP model (U.S. EPA, 1997, 2003); 3) the Industrial Subtitle D Facility Study (also called the Subtitle D survey), conducted by

the U.S. EPA OSW which provides a statistically valid set of site-specific areas, volumes and locations for industrial Subtitle D landfills, waste piles, and land application units around the country; and 4) the EPA's recent 5-year nationwide study of nonhazardous (subtitle D) industrial surface impoundments (the SI Study) which provides a statistically valid set of site-specific impoundment characteristics, including impoundment location, area, ponding depth, and operational life.

The HGDB, developed by Rice University for American Petroleum Institute in 1989, provides site specific data on ground-water parameters (aquifer thickness, depth to ground water, hydraulic gradient and hydraulic conductivity) collected by independent investigators for approximately 400 hazardous waste sites throughout the U.S. These site-specific data were then regrouped into 13 hydrogeologic environments, based on the USGS classification of aquifer regions (Heath, 1984). The result is a database of aquifer types, with each aquifer type consisting of an empirical distribution of values for each of the four aquifer parameters.

Infiltration and recharge rates for use in EPACMTP modeling applications have been estimated for selected soil types at cities around the country through the use of the HELP water-balance model. Using the Soil Conservation Service's (SCS) county-by-county soil mapping database, three soil textures were defined: coarse-; medium-; and fine-grained soils. Using National Oceanic and Atmospheric Administration (NOAA) data on precipitation and evaporation rates in the United States, 102 cities were selected as climatic centers for the HELP model. For each selected city, historical climatic data were used to develop an ambient regional recharge rate as a function of site location and soil type; likewise, infiltration rates were developed for each type of WMU, as a function of site location, liner type (if any) and soil type.

We have tabulated the results of the Industrial Subtitle D Facility Study (U.S. EPA, 1986) into a nationwide database of waste management unit sites for use in probabilistic EPACMTP modeling analyses of landfills, waste piles and land application units. The original survey provides a set of observations of site-specific areas, volumes and locations for Industrial Subtitle D landfill, waste pile, and land application facilities across the U.S. Although surface impoundments were included in the Industrial Subtitle D Facility Study, EPA has adopted the results of the more recent Surface Impoundment (SI) Study as the data source for the database of surface impoundment sites. The SI Study provided data on impoundment locations, area, operating depths (depth of ponding in the impoundment), depth of the SI base below the ground surface, operational life of the impoundment, and proximity of the impoundment to a surface water body. Since the Subtitle D survey and the SI Study include only facility-specific data, linkages to the other two data sources (HGDB and the HELP-modeled climatic database) are used to generate the additional input parameters required to perform the ground-water fate and transport modeling for each site. That is, for use in EPACMTP modeling analyses, the modelers classified each site in the Subtitle D survey and the SI Study databases according to the type of aquifer underlying the site and the closest climate center used in the HELP modeling in order to provide links to the hydrogeologic and climatic databases. Details of the analysis and screening of SI facilities and units are presented in the EPA SI Study report (U.S. EPA, 2001). Data on various types of waste management

units for use in EPACMTP are provided in the *EPACMTP Parameters/Data Background Document* (U.S. EPA, 2003).

5.5.2 Regional Site-Based Monte-Carlo Procedure

Fundamentally, the approach used for a site-based Monte-Carlo analysis consists of conducting the modeling analysis for the sites in a waste site database (either the Subtitle D survey or the SI Study) on the assumption that these sites are an adequate representation of the universe of possible waste sites in the U.S. The actual procedure of the Monte-Carlo simulation is summarized in a number of steps below. These steps describe the general procedure used at the time of preparation of this document and reflect the currently available databases commonly used by EPA. Some of the specifics may vary for individual projects as EPA updates its databases, but the steps below will still illustrate the essence of the methodology:

STEP 1: Select a Waste Site

The first step involves selecting a site, at random, from the list of waste sites. The data set is treated as an empirical distribution. In most instances, each site will have an equal probability of occurrence, although it is possible to vary the probability so that some sites have a greater likelihood of being selected than others. When the model selects a site, the data read into EPACMTP include the appropriate characteristics for that site, including unit area, unit depth, an index that specifies the nearest climate center, and an index that specifies the underlying aquifer type.

STEP 2: Generate Recharge and Infiltration for the Selected Waste Site

The soil type at the chosen waste site (and cover type for landfills) and the specified liner scenario (no liner/in-situ soil, single clay liner, or composite liner) are then used along with the climate center index to determine the appropriate values for recharge and infiltration at the site by querying the database of HELP-modeled recharge and infiltration rates. The specific soil (and landfill cover type) can be specified individually for each waste unit or as a probability distribution.

STEP 3: Generate Hydrogeologic Variables for Selected Site in the Industrial Subtitle D Facility Study

As explained above, given the resolution of available hydrogeological databases and acknowledging the uncertainty in the effective local values of hydrogeological site characteristics, the regional site-base approach is generally implemented by using the aquifer type assigned to the chosen site in the WMU database. The input values for the hydrogeologic parameters for the chosen site are then determined from the probability distributions that define the corresponding aquifer type. That is, a correlated set of hydrogeologic parameter values is randomly chosen from among those available in the hydrogeologic database for the chosen aquifer type. If the selected ground-water parameter set is missing any values, a joint distribution of the parameters (derived for each environment) is used to fill in the missing values. The details of this procedure are presented in Section 5.5.3.

STEP 4: Generate Remaining Parameters for the Selected Waste Site

The remaining parameters for the waste site that are not assigned as site-related (e.g., x, y, and z coordinates of the receptor well) are generated by using one of the probability distributions described in Section 5.4. Any derived parameters are calculated as described in Section 5.4.10.

STEP 5: Calculate the Predicted Receptor Well Concentration Value for the Selected Waste Site

Given the complete set of input parameter values generated in the previous four steps, and the chemical-specific characteristics (e.g., leachate concentration, adsorption coefficient and exponent, and degradation rate), the EPACMTP flow and transport modules are executed to compute the receptor well concentration value for this Monte-Carlo realization.

STEP 6: Repeat Steps 1 Through 5 a Specified Number of Times, and Estimate the National Distribution of Receptor Well Concentrations

After Step 5, the receptor well concentration value for a specific realization is obtained. The process is then repeated as many times as is specified by the user. The result of the Monte-Carlo modeling analysis is a receptor well concentration for each model realization; these results represent the nationwide distribution of drinking water exposure concentrations. In most cases, the number of Monte-Carlo realizations will be much greater than the number of sites used in the regional, site-based analysis, and because the selection of sites as described in Step 1 above is random, each site is expected to be picked more than once. However, because there is an additional random component to the process of assigning values to all of the EPACMTP parameters (e.g., well location), repeated selection of the same waste site in the Monte-Carlo process generally will not result in the same predicted receptor well exposure concentration.

5.5.3 Methodology for Generating Missing Data Values

Below is a step-by-step presentation of the methodology used, within the framework of the site-based Monte-Carlo approach, to generate missing parameter values, based on the statistical correlation between parameters with missing values and model input parameters whose values are known.

For each parameter of interest, for example, hydraulic conductivity, we have a set of known (observed) values, but also a number of missing values. For the known values we also have corresponding values of related parameters, e.g., hydraulic gradient and saturated thickness. From this information we constructed a covariance matrix that expresses the statistical relationships among all parameters. Given this covariance matrix and values for one or more parameters it is possible to estimate missing values for the parameter of interest. For instance, we can estimate the value of hydraulic conductivity given values for hydraulic gradient and/or saturated thickness. The methodology described here is applicable to parameters with multi-variate normal (Gaussian) distributions. Consequently, if the actual

parameters have non-Gaussian distributions, each must first be transformed to a normal distribution. If applicable, any statistically generated values must also be back-transformed to obtain the parameter value in the original, untransformed space. The parameter covariance matrix is required also. The covariance matrix is calculated from the transformed variables.

The algorithm to generate missing values is described using vector notation. In the notation below, the superscript ^{T} denotes the transpose operator, i.e. switching of rows and columns, and superscript $^{-1}$ denotes an inverse matrix.

1. The process begins with a set of parameters transformed to a normal distribution if necessary. The set can be expressed as a data vector, X:
   X = (n × 1) data vector of normally distributed, correlated parameters
   = $(X_1,..,.. X_n)^T$

2. The second step is to create a vector Y by partitioning the data vector x so that the first p elements of Y correspond to the missing values of X, and the remaining q elements correspond to the observed values of X:

   Y = $(Y_1, Y_2)^T$
   Y_1 = (p × 1) vector of missing values of X
   Y_2 = (q × 1) vector of observed values of X

3. As stated above, Y consists of normally distributed parameters and known correlations (covariance) among parameters. The parameter vector Y can therefore be expressed statistically in terms of a multivariate normal distribution characterized by a vector of mean values, m, and a covariance matrix V:

$$Y \sim N_n(\overline{m},V) \tag{5.9}$$

where

$N_n()$ = n-variate normal distribution with mean vector m and covariance matrix V
$\overline{m}$ = $(m_1, m_2)^T$
n = $p + q$
m_1 = (p × 1) mean vector of missing values
m_2 = (q × 1) mean vector of observed values
V = (n × n) covariance matrix

Given that Y is composed of unknown (Y_1) and known (Y_2) values, its multivariate statistical representation can also be portioned accordingly. Given Y_2, the conditional distribution of Y_1 which is a multivariate normal distribution with (p × 1) mean vector $m_{1.2}$ and (p × p) covariance matrix $V_{1.2}$ is:

$$\boldsymbol{Y_1} \mid \boldsymbol{Y_2} \sim \boldsymbol{N_n}\,(\boldsymbol{m_{1.2}}, \boldsymbol{V_{1.2}}) \qquad (5.10a)$$

$$\boldsymbol{m_{1.2}} = \boldsymbol{m_1} + \boldsymbol{V_{12}}\,\boldsymbol{V_{22}^{-1}}\,(\boldsymbol{Y_2} - \boldsymbol{m_2}) \qquad (5.10b)$$

$$\boldsymbol{V_{1.2}} = \boldsymbol{V_{11}} - \boldsymbol{V_{12}}\,\boldsymbol{V_{22}}\,\boldsymbol{V_{21}} \qquad (5.10c)$$

where

Y_1 = vector of missing values
Y_2 = vector of observed values
N_n = n-variate normal distribution with mean vector m and covariance matrix V
$m_{1.2}$ = (p × 1) mean vector of $\boldsymbol{Y_1}$ conditioned by $\boldsymbol{Y_2}$
$V_{1.2}$ = ($p \times p$) covariance matrix of Y, conditioned by Y_2
m_1 = vector of means of missing values
$\boldsymbol{V_{22}^{-1}}$ = inverse of V_{22}
m_2 = vector of means of observed values
$\boldsymbol{V_{11}}$ = ($p \times p$) upper left partition of V
$\boldsymbol{V_{12}}$ = ($p \times q$) upper right partition of V
$\boldsymbol{V_{21}}$ = ($q \times p$) lower left partition of V
$\boldsymbol{V_{22}}$ = ($q \times q$) lower right partition of V

4. The steps above now lead to the following equation for estimating unknown values of Y. Given the observed vector Y_2, a prediction of the missing vector $Y_{1.2}$, is generated by:

$$\boldsymbol{Y_{1.2}} = \boldsymbol{m_{1.2}} + \boldsymbol{L^*}(\boldsymbol{D^*})^{½}\,\boldsymbol{u} \qquad (5.11a)$$

where

$\boldsymbol{Y_{1.2}}$ = prediction of the missing vector $\boldsymbol{Y_1}$
$\boldsymbol{m_{1.2}}$ = (p × 1) mean vector of $\boldsymbol{Y_1}$ conditioned by $\boldsymbol{Y_2}$
L^* = (p × p) matrix of the eigenvectors of $\boldsymbol{V_{1.2}}$

u = (p × 1) vector of independent and identically distributed standard normal random variables

D^* = (p × p) diagonal matrix consisting of the square root of the eigenvalues of $V_{1.2}$, so that

$$V_{1.2} = B^* B^{*T} \tag{5.11b}$$

where

$V_{1.2}$ = ($p \times q$) upper right partition of V

B^* = (p × p) matrix of square root of $V_{1.2} = L^*(D^*)^{½}$

B^{*T} = transpose of B^*

An inspection of equation (5.11 a) shows that missing values are estimated from their mean, $m_{1.2}$, plus a contribution from the covariance with other parameters, expressed by $L^*(D^*)^{½}$. This contribution also includes a random factor, u. If the correlation between the parameter whose value is unknown and related parameters is weak, then the second term on the right-hand side of (5.11 a) will tend to be close to zero, and the estimated value will be close to the mean. Conversely, a strong correlation means that $L^*(D^*)^{½}$ will have a higher value and this will allow the estimated value for Y to be more different from the mean. The incorporation of u means that there always is a random component to the estimate, except in the case of zero correlation when $L^*(D^*)^{½}$ is exactly zero.

5.6 INTERPRETING A MONTE-CARLO MODELING ANALYSIS

The result of a Monte-Carlo simulation is a sequence of receptor well concentration values. Each value corresponds to one Monte-Carlo realization. Collectively, they represent the range of possible outcomes for the EPACMTP modeling scenario of interest based on the probability distributions assigned to each of the EPACMTP input parameters. The Monte-Carlo outputs are best analyzed and interpreted in terms of probability. For ease of interpretation, it is often convenient to normalize the computed receptor well concentrations to the (initial) value of the leachate concentration infiltrating to the subsurface from the base of the waste unit:

$$\overline{C_r} = \frac{C_{rwell}}{C_L} \tag{5.12}$$

where

$\overline{C_r}$ = relative concentration at receptor well (dimensionless)

C_{rwell} = constituent concentration at receptor well (mg/L) (instantaneous or time-averaged)

C_L = leachate concentration (mg/L)

If the modeling scenario includes a time-varying leachate concentration, as in the case of a depleting landfill, the value of the initial leachate concentration is used in Equation 5.12. $\bar{C}_r$ is a dimensionless quantity, with a value between zero and one. It is called the normalized or relative concentration. The reduction in concentration between the leachate concentration which enters the subsurface and the eventual concentration predicted to occur at the receptor well is a result of dilution and attenuation processes which occur during the transport of the constituent through soil and ground water. A convenient way to express the aggregate effects of all fate and transport processes simulated by EPACMTP is in terms of the Dilution-Attenuation Factor (DAF) which is defined as:

$$DAF = \frac{C_L}{C_{rwell}} = \frac{1}{\bar{C}_r} \qquad (5.13)$$

where

DAF = dilution-attenuation factor

The DAF is a dimensionless quantity, the value of which can vary from one (1) to infinity. A DAF value of 1 corresponds to a relative receptor well concentration of one. This situation means that the exposure concentration at the receptor well is the same as the leachate concentration that enters the subsurface from the modeled waste management unit, and no dilution or attenuation occurs along the subsurface pathway. Conversely, if the contaminant plume does not reach the receptor well at all, the receptor well concentration will be zero and the corresponding DAF will approach infinity.

For organic constituents, the fate and transport equations solved by EPACMTP are linear, which means that the magnitude of the predicted ground-water well concentration is linearly proportional to the value of the leachate concentration. In other words, for organics, a doubling of the EPACMTP input value of leachate concentration would result in a doubling of the predicted ground-water well concentration, as long as all other model parameters stay the same. Equation 5.13 is applicable to chemicals with both linear and non-linear sorption isotherms. For chemicals with linear sorption isotherms, their DAFs are not dependent on leachate concentrations. In other words, once the DAF for a chemical with a linear isotherm has been determined, it can be used to determine C_{rwell} regardless of the value of C_L. On the other hand, DAFs for constituents whose geochemical behavior is characterized by nonlinear sorption isotherms (i.e., metals) are C_L-specific, and C_{rwell} is not linearly related to C_L. For this reason, the DAF is a less useful concept for describing the transport behavior of metals.

Conceptually, each Monte-Carlo realization represents one possible real-world outcome, and each realization has an equal probability of occurrence. A Monte-Carlo simulation will result in a distribution of predicted receptor well concentrations, and through a post-processing step the EPACMTP user can obtain the probability distribution of the expected receptor well exposure concentrations – or DAFs – by

constructing a simple frequency histogram of the Monte-Carlo modeling results. An example is shown in Figure 5.2.

Figure 5.2 shows the frequency distribution of the normalized receptor well concentrations obtained in a EPACMTP Monte-Carlo analysis. Frequency is expressed on a normalized scale from 0 to 100. Although this is a fictitious example, it does illustrate a number of typical features of a Monte-Carlo simulation. One of these key features is that many of the Monte-Carlo realizations result in very low concentrations at the receptor well. In the example shown, the receptor well concentration is 1/50th or less of the leachate concentration in more than 80% of the cases. Correspondingly, there are relatively few occurrences of high normalized concentration values. In showing these features, Figure 5.2 also illustrates that a regular frequency histogram is not the most convenient way to present the results. A more useful way to do this is to present the Monte-Carlo results in the form of a cumulative frequency graph, otherwise known as a Cumulative Distribution Function (CDF). Figure 5.3 presents the data from Figure 5.2 as a CDF.

Figure 5.3 shows the cumulative frequency expressed as a percentile. EPA often summarizes a Monte-Carlo analysis in terms of specific percentile values of the CDF of the normalized receptor well concentration, or its corollary, the DAF. In the example shown in Figure 5.3, the 90th percentile of the concentration CDF corresponds to a normalized receptor well concentration of about 0.1. This means that in 90% of the cases, the receptor well concentration is one-tenth or less of the leachate concentration at the waste management unit. An equivalent statement is that the 10th percentile of DAF is 10 (the reciprocal of a normalized receptor well concentration value of 0.1)

There are several ways to summarize the results of the Monte-Carlo analysis process. For instance, the resulting distribution of receptor well concentrations can be analyzed to identify the percentage of realizations that produce a receptor well concentration above or below a specified ground-water reference concentration (such as a Maximum Contaminant Level (MCL)). Alternatively, using the DAF, the input leachate concentration can be scaled to calculate the leachate concentration threshold value – that is, the maximum allowable leachate concentration that results in a predicted receptor well concentration being less than the ground-water reference concentration (such as an MCL) in a defined percentage of the model realizations. Assuming that the Monte Carlo modeling process indeed captures the range of variability and uncertainty encountered at actual waste sites across the United States, the Monte-Carlo results indicate the fraction of sites for which expected receptor well concentrations are less (or DAFs higher) than a particular threshold value. This, in turn, provides the basis for developing regulatory leachate and/or waste concentration threshold values and determining appropriate waste management requirements to ensure compliance with risk-based or other ground-water quality criteria.

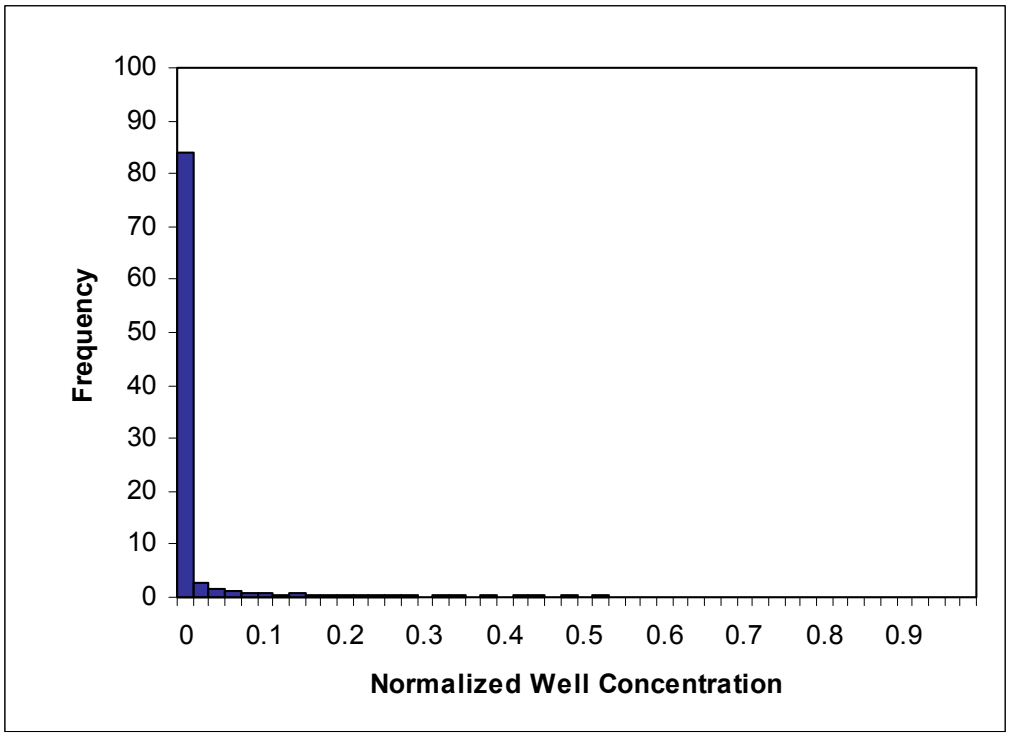

Figure 5.2 **Frequency distribution of normalized receptor well concentrations.**

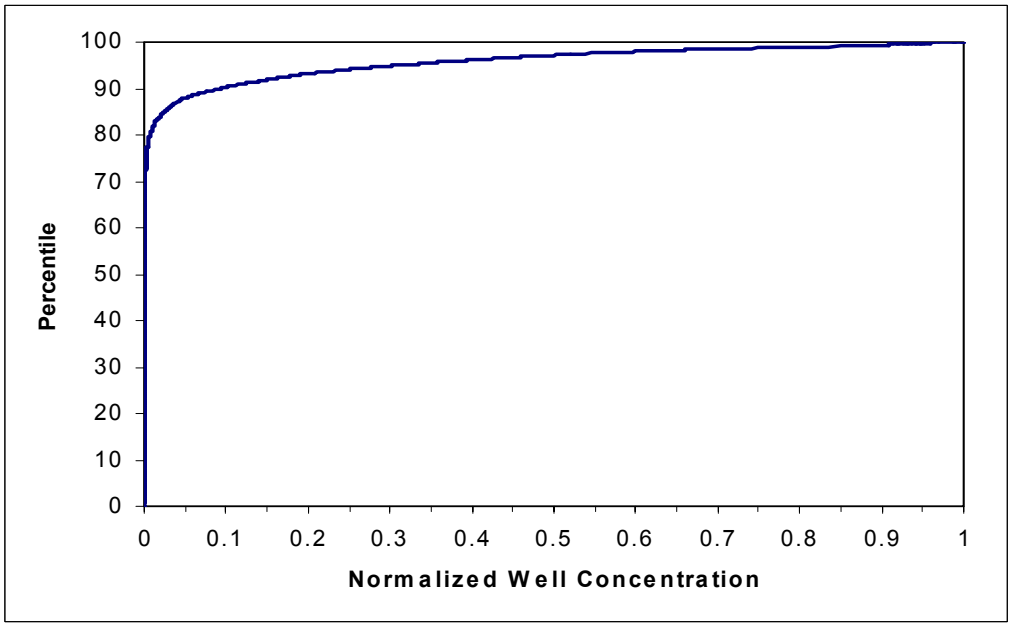

Figure 5.3 **Cumulative Distribution Function of Normalized Receptor Well Concentration**.

For instance, consider the following question: "What is the maximum allowable leachate concentration for chemical x in this waste stream that is protective of human health in at least 90% of the cases ?" This question can be answered in the following manner. For exposure to ground water, the human health standard (also called the Reference Ground-water Concentration or RGC) can be expressed in terms of a ground-water exposure concentration value (e.g., health-based number (HBN) or maximum contaminant level (MCL)) at a well intake point, corresponding to an acceptable risk level, say 10^{-6} cancer risk. The corresponding maximum allowable leachate concentration can then be back-calculated using the Monte-Carlo modeling results as:

$$C_L^{max} = DAF_{10} \times HBN \qquad (5.14)$$

where

C_L^{max} = maximum allowable leachate concentration (mg/L)
DAF_{10} = 10th percentile value of DAF (which corresponds to the 90th percentile of relative concentration) (dimensionless)
HBN = Health-Based Number, which is a ground-water exposure concentration corresponding to a defined risk level (mg/L)

Using the DAF values from the example presented in this section and a protection level of 90%, the maximum allowable leachate concentration would be 10 times the health-based ground-water concentration threshold, reflecting the fact that we expect the dilution and attenuation during ground-water transport to be a factor of 10 or greater in at least 90% of the modeled cases.

5.7 REQUIRED NUMBER OF MONTE-CARLO REALIZATIONS

It is inherent in the random sampling approach of a Monte-Carlo analysis that the modeling outcome depends on the number of realizations. For instance, the estimate of the 90th percentile predicted ground-water concentration will likely be different if we calculate it from 100 realizations, as compared to 1,000 realizations. In using a Monte-Carlo modeling approach, a higher number of realizations usually leads to a more convergent and reliable result. Results are said to be converged if the estimate of a particular percentile value does not change significantly if additional Monte-Carlo simulations are performed. However, it is not generally possible to determine beforehand how many realizations are needed to achieve a specified degree of convergence since the value can be highly dependent on parameter distributions.

EPA conducted a bootstrap analysis for the EPACMTP model to evaluate how convergence improves with increasing number of realizations. A bootstrap analysis is a technique of replicated re-sampling (usually by a computer) of an original data set for estimating standard errors, biases, confidence intervals, or other measures of statistical accuracy. Bootstrap analysis can automatically produce accuracy estimates in almost any situation without requiring subjective statistical assumptions

about the original distribution. The EPACMTP modeling scenario represented in EPA's bootstrap analysis was that of a continuous source, landfill disposal scenario in which the "true" 10th percentile DAF was 10. The results of the bootstrap analysis are summarized in Table 5.3. These results show that, with 10,000 realizations, the expected value of the 10th percentile DAF was 10 with a 95 percent confidence interval of 10 ± 0.7. Decreasing the number of realizations to 5,000 increased the confidence interval to 10 ± 1.0.

Table 5.3 Relationship between confidence interval and number of Monte-Carlo realizations.

Number of Realizations	97.5 Percent Confidence Interval	
	Lower Limit of DAF_{10}	Upper Limit of DAF_{10}
1,000	7.52	12.74
2,000	8.36	11.60
5,000	9.02	11.06
10,000	9.31	10.74
20,000	9.51	10.46
30,000	9.63	10.40

This bootstrap analysis illustrates the relatively slow decrease in the prediction error as the number of Monte-Carlo realizations is increased.

EPA has adopted 10,000 model realizations in recent EPACMTP modeling applications. The actual number of realizations adopted for regulatory analyses by EPA is a balance between the desire for optimal convergence and practical constraints of resources and time needed to perform large numbers of computer analyses, as well as considering the relative benefit of increasing the number of Monte-Carlo realizations against other inherent sources of uncertainty. The diminishing benefit of increasing the number of Monte-Carlo realizations is illustrated in Figure 5.4 in this figure. The relative prediction error from the bootstrap analysis is plotted against the number of realizations. The relative error here is defined as 95 percent confidence interval (difference between upper and lower confidence limits in Table 5.3), divided by the number of Monte-Carlo realizations. The results in the figure are multiplied by a scaling factor of 1,000 for presentation purposes. This figure illustrates that when the number of Monte-Carlo realizations is fairly small, that is on the order of 1,000, increasing this number can significantly reduce the prediction error. However, this benefit diminishes when the number of realizations is already on the order of 10,000 or greater.

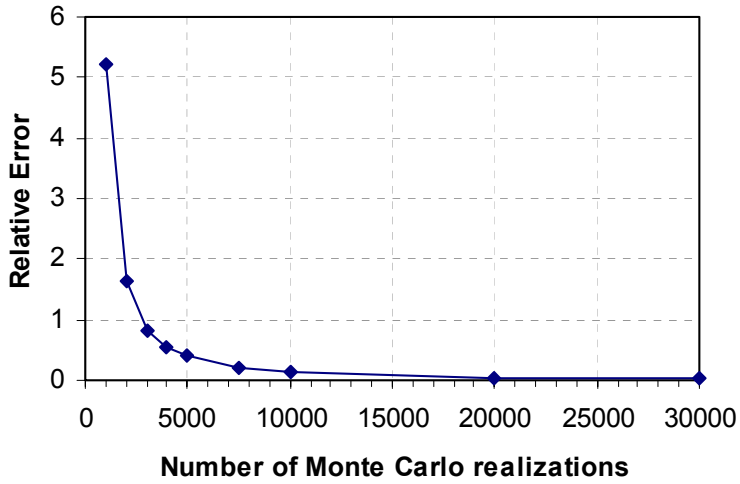

Figure 5.4 Relative Monte-Carlo Prediction Error.

6.0 REFERENCES

Bear, J., 1972. *Dynamics of Fluids in Porous Media*. American Elsevier, New York, NY.

Bear, J., 1979. *Hydraulics of Groundwater.* McGraw Hill, New York, NY.

Bonaparte, R., J.P. Giroud, and B.A. Cross, 1989. Rates of Leakage through Landfill liners. *Geosynthetics 1989 Conference*, San Diego, California.

Burnett, R.D. and E.O. Frind, 1987. Simulation of contaminant transport in three dimensions. 2. Dimensionality effects. *Water Resources Research* 23(4): 695-705.

Carsel, Robert F., and R.S. Parrish, 1988. A method for developing joint probability distributions of soil-water retention characteristics. *Water Resources Research* 5: 755-700.

Carsel, Robert F., Rudolph S. Parrish, Russell L. Jones, James L. Hansen, and Richard L. Lamb, 1988. Characterizing the uncertainty of pesticide leaching in agricultural soils. *Journal of Contaminant Hydrology* 2: 111-124.

Carslaw, H.S., and Jaeger, J.C., 1959. *Conduction of Heat in Solids.* 2nd Ed., Oxford University Press.

CRC, 1981. *Handbook of Chemistry and Physics.* 62nd Ed. CRC Press.

Davis, S.N., 1969. Porosity and permeability of natural materials. In: *Flow Through Porous Media*, R.J.M. de Wiest, Editor, Academic Press, NY.

de Hoog, F.R., J.H. Knight, and A.N. Stokes, 1982. An improved method for numerical inversion of Laplace transforms. *SIAM J. Sci. Stat. Comput.* 3(3): 357-366.

de Marsily, G., 1986. *Quantitative Hydrology*, Academic Press, Orlando, FL.

Domenico, P.A. and V.V Palciauskas., 1982. Alternative boundaries in solid waste management, *Ground Water* 20: 301-311.

Domenico, P.A., and G.A. Robbins, 1985. A new method for contaminant plume analysis, *Ground Water* 25(6): 733-739.

Enfield, C.G. et al., 1982. Approximating pollutant transport to groundwater. *Ground Water* 10(6): 711-722.

Freeze, R. Allen and John A. Cherry, 1979. *Groundwater*. Prentice-Hall, Englewood Cliffs, NJ.

Gelhar, L.W., A. Mantoglou, C. Weltey, and K. R. Rehfeldt, 1985. *A review of field-scale physical solute transport processes in saturated and unsaturated porous media*, Electric Power Research Institute Report EA-4190, Project 2485-5, Palo Alto, CA, 1985.

Gelhar, L.W., 1986. Personal communication

Gelhar, L.W., C. Welty, K.R. Rehfeldt, 1992. A critical review of data on field-scale dispersion in aquifers. *Water Resources Research* 28(7): 1955-1974.

Harleman, D.R.F., and R.R. Rumer, 1963. Longitudinal and Lateral Dispersion in Isotropic Porous Medium, *J. Fluid Mech.* 16(3): 385-394.

Heath, R.C., 1984. National Water Summary 1984. Section entitled *State Summaries of Groundwater Resources*. USGS Water-Supply Paper 2275.

Huyakorn, P.S., and K. Nilkuha, 1979. Solution of Transient Transport Equation Using an Upstream Finite Element Scheme. *Appl. Math. Modelling* 3: 7-17.

Huyakorn, P.S., and G.F. Pinder, 1983. *Computational Methods in Subsurface Flow*, Academic Press.

Huyakorn, P.S., B.G. Jones, and P.F. Andersen, 1986. Finite Element Algorithms for Simulating Three-Dimensional Groundwater Flow and Solute Transport in Multilayer Systems. *Water Resources Research* 22(3):361-374.

Huyakorn, P.S., and J.E. Buckley, 1987. *VADOFT: Finite Element Code for Simulating One-Dimensional Flow and Solute Transport in the Vadose-zone*. Technical Report Prepared for the U.S. Environmental Protection Agency, Athens, GA.

Karickhoff, S.W., 1985. *Sorption protocol for evaluation of OSW chemicals.* U.S. EPA, Athens Environmental Research laboratory, Athens, GA.

Lambe, T.W., and Whitman, R.V., 1969. *Soil Mechanics.* John Wiley and Sons.

MacDonald, J.R., 1964. Accelerated Convergence, Divergence, Iteration, Extrapolation, and Curve Fitting. *J. Appl. Phys.* 10: 3034-3041.

McGrath, E.J. and D.C. Irving, 1973. *Techniques for Efficient Monte Carlo Simulation.* Technical report prepared for Department of the Navy, Office of Naval Research, Arlington, VA.

McWorter, D.B., and D.K. Sunada, 1977. *Groundwater Hydrology and Hydraulics*. Water Resources Publications, Fort Collins, CO.

Millington, R.J. and J.P. Quirk, 1961. Permeability of porous solids, *Trans. Faraday Soc.* 57: 1200-1207.

Newell, Charles J., Loren P. Hopkins, and Philip B. Bedient, 1989. *Hydrogeologic Database for Ground Water Modeling.* API Publication No. 4476, American Petroleum Institute, Washington, DC.

Newell, Charles J., Loren P. Hopkins, and Philip B. Bedient, 1990. Hydrogeologic Database for Ground Water Modeling. *Ground Water* 28(5): 703 - 714.

Rollin, A.L., M. Marcotte, T. Jacquelin, and L. Chaput, 1999. Leak Location in Exposed Geomembrane Liners using an Electrical Leak Detection Technique. *GeoSynthetics '99: Specifying Geosynthetics and Developing Design Details* 2: 615-626.

Schroeder, Paul, Cheryl Lloyd, Paul Zappi, and Nadim Aziz, 1994a. *The Hydrologic Evaluation of Landfill Performance (HELP) Model: User's Guide for Version 3.* EPA/600/R-94/168a. U.S. EPA Risk Reduction Engineering Laboratory, Cincinnati, OH.

Schroeder, Paul, Cheryl Lloyd, Paul Zappi, and Nadim Aziz, 1994b. *The Hydrologic Evaluation of Landfill Performance (HELP) Model: Engineering Documentation for Version 3.* EPA/600/R-94/168b. U.S. EPA Risk Reduction Engineering Laboratory, Cincinnati, OH.

Science Advisory Board, 1988. *Review of the Unsaturated Zone Code for the OSW Fate and Transport Model.* Report of the Unsaturated Zone Code Subcommittee. SAB-EEC-88-030. June, 1988.

Science Advisory Board, 1990. *Review of the CANSAZ Flow and Transport Model for Use in EPACMS.* Report of the Saturated Zone Model Subcommittee. EPA-SAB-EEC-90-009. March, 1990.

Science Advisory Board, 1995. An SAB Report: *Review of EPA's Composite Model for Leachate Migration with Transformation Product-EPACMTP.* EPA-SAB-EEC-95-110. August, 1998.

Sudicky, E.A, 1989. The Laplace Transform Galerkin Technique: A Time-Continuous finite Element theory and Application to Mass Transport in Groundwater. *Water Resources Research* 25 (8): 1833-1846.

Sudicky, E.A., J.B. Kool and P.S. Huyakorn, 1990. *CANSAZ: Combined Analytical-Numerical Model for Simulating Flow and Contaminant Transport in the Saturated Zone. Version 2.0 with Nonlinear Adsorption and Chain Decay Reactions.* Technical Report Prepared for U.S. EPA, Office of Solid Waste, Washington, DC, 20460.

Sudicky, E.A., and R.G. McLaren, 1992. The Laplace Transform Galerkin Technique for Large-Scale Simulation of Mass Transport in discretely-Fractured Porous Formations. *Water Resources Research* 28 (2): 499-514.

Todd, D.K., 1959. *Ground Water Hydrology*. John Wiley, New York.

U.S. EPA, 1986. Industrial Subtitle D Facility Study (Telephone Survey), U.S. Environmental Protection Agency, October 20, 1986.

U.S. EPA , 1989. *Unsaturated Zone Flow and Transport Module: Version 2.0 with Nonlinear Sorption and Chain Decay Reactions.* U.S. Environmental Protection Agency, Office of Solid Waste, Washington, DC, 20460.

U.S. EPA, 1990. *Background Document for EPA's Composite Model for Landfills (EPACML).* U.S. Environmental Protection Agency, Office of Solid Waste, Washington, D.C. 20460.

U.S. EPA, 1991. *Standard Default Exposure Factors*, OSWER Directive 9285.6-03, U.S. Environmental Protection Agency Office of Emergency and Remedial Response, Washington, DC.

U.S. EPA, 1992. *Guidelines for Exposure Assessment.* EPA-600Z-92-001. Office of Research and Development, Risk Assessment Forum, Washington, DC. May.

U.S. EPA, 1996a. *EPA's Composite Model for Leachate Migration with Transformation Products: (EPACMTP) Technical Background Document.* U.S. Environmental Protection Agency, Office of Solid Waste, Washington, D.C. 20460.

U.S. EPA, 1996b. *EPA's Composite Model for Leachate Migration with Transformation Products: (EPACMTP) Background Document for Finite Source Methodology for Chemicals with Transformation Products.* U.S. Environmental Protection Agency, Office of Solid Waste, Washington, D.C. 20460.

U.S. EPA, 1996c. *EPACMTP Sensitivity Analyses.* U.S. Environmental Protection Agency, Washington, D.C. 20460.

U.S. EPA, 1997. *Analysis of EPA's Industrial Subtitle D Databases used in Groundwater Pathway Analysis of the Hazardous Waste Identification Rule (HWIR).* U.S. Environmental Protection Agency, Office of Solid Waste, Washington, DC.

U.S. EPA, 1999. *EPA's Composite Model for Leachate Migration with Transformation Products, Background Document for EPACMTP: Metals Transport in the Subsurface, Volume 1: Methodology.* U.S. Environmental Protection Agency, Office of Solid Waste, Washington, D.C. 20460.

U.S. EPA, 2001. *Industrial Surface Impoundments in the United States.* U.S. Environmental Protection Agency, Office of Solid Waste, Washington, DC 20460. U.S. EPA 530-R-01-005.

U.S. EPA, 2003. *EPA's Composite Model for Leachate Migration with Transformation Products (EPACMTP): Parameters/Data Background Document.* U.S. Environmental Protection Agency, Office of Solid Waste, Washington, D.C. 20460. U.S. EPA 530-R-03-003.

van Genuchten, M.Th., 1980. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils, *Soil Sci. Soc. J.* 44: 892-898.

This page was intentionally left blank.